

```
In [1]: import urllib.request
        from bs4 import BeautifulSoup
        from imdb import IMDb
```

```
In [57]: def get_film_info(url):
        result = {}
        s = requests.Session()
        s.headers.update({
            'User-Agent': 'Mozilla/5.0 (X11; Linux x86_64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/81.0.404
            'Cookie': 'profile_button=votes; fuid01=5b54ad0b691302bf.4-UihjJRc4aQbk37oRHyoitU9VI8iVEYjFFph0YWe9k6
        })
        r = s.get(url)
        proxies = {
            "https": "https://82.119.170.106:8080",
            "http": "http://82.119.170.106:8080"
        }
        film = BeautifulSoup(r.text, 'lxml')
        result['name'] = film.find(class_='moviename-title-wrapper').text

        with open('films/' + result['name'].replace(' ', '_') + '.html', 'w') as output_file:
            output_file.write(r.text)

        result['rating'] = film.find(class_='rating_ball').text
        result['views'] = film.find(class_='ratingCount').text

        table = film.find('table', class_='info')
        for tr in table.find_all('tr'):
            td1, td2 = tr.find_all('td')
            if td1.text in ['год', 'страна', 'режиссер']:
                result[td1.text] = td2.find('a').text
            if td1.text == 'время':
                result[td1.text] = td2.text
        #     time.sleep(1)
        return result
```

```
In [3]: get_film_info('https://www.kinopoisk.ru/film/524/')
```

```
Out[3]: {'name': 'Дилижанс',  
        'rating': '7.524',  
        'views': '3452',  
        'время': '96 мин. / 01:36',  
        'год': '1939',  
        'режиссер': 'Джон Форд',  
        'страна': 'США'}
```

```
In [4]: def get_film_links():  
        start_url = 'https://www.kinopoisk.ru/lists/top250/?page={}&tab=all'  
        result = []  
        s = requests.Session()  
        s.headers.update({  
            'User-Agent': 'Mozilla/5.0 (Macintosh; Intel Mac OS X 10.9; rv:45.0) Gecko/20100101 Firefox/45.0'  
        })  
        for i in range(1,10):  
            r = s.get(start_url.format(i))  
            page = BeautifulSoup(r.text, 'lxml')  
            films = page.find_all(class_='selection-film-item-meta selection-film-item-meta_theme_desktop')  
            for film in films:  
                result.append('https://www.kinopoisk.ru' + film.find('a').get('href'))  
        return result
```

```
In [5]: links = get_film_links()
```

```
In [6]: len(links)  
        # with open('kinopoisk_links.txt', 'w') as file:  
        #     for link in links:  
        #         file.write(link + '\n')
```

```
Out[6]: 220
```

```
In [9]: data = []
```

```
In [58]: for i in range(221):
         data.append(get_film_info(links[i]))
```

```
-----
IndexError                                Traceback (most recent call last)
<ipython-input-58-44be3682c1a2> in <module>()
      1 for i in range(213, 221):
----> 2     data.append(get_film_info(links[i]))

IndexError: list index out of range
```

```
In [68]: len(data)
```

```
Out[68]: 219
```

```
In [61]: import pandas as pd
```

```
In [64]: table = pd.DataFrame(data)
         table.head()
```

```
Out[64]:
```

	name	rating	views	время	год	режиссер	страна
0	Побег из Шоушенка	9.111	693 385	142 мин. / 02:22	1994	Фрэнк Дарабонт	США
1	Зеленая миля	9.062	594 470	189 мин. / 03:09	1999	Фрэнк Дарабонт	США
2	Форрест Гамп	8.913	561 423	142 мин. / 02:22	1994	Роберт Земекис	США
3	Список Шиндлера	8.818	323 887	195 мин. / 03:15	1993	Стивен Спилберг	США
4	1+1	8.807	1 006 352	112 мин. / 01:52	2011	Оливье Накаш	Франция

```
In [67]: table.to_csv('data.csv', index=False)
```

```
In [ ]:
```